EDORA - An expert system for assistance in biological modeling

Leila Alem

INRIA

UNITÉ DE RECHERCHE
INRIA-SOPHIA ANTIPOLIS

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France
Tél. (1) 39 63 55 11

Rapports Techniques

N° 99

EDORA - AN EXPERT SYSTEM FOR ASSISTANCE IN BIOLOGICAL MODELING

Programme 5

Leila ALEM

OCTOBRE 1988

RT-099

Leila ALEM

INRIA Sophia-Antipolis,
Route des Lucioles
06560 Valbonne, FRANCE.

EDORA : AN EXPERT SYSTEM FOR ASSISTANCE IN BIOLOGICAL MODELING

EDORA : UN SYSTEME EXPERT D'AIDE A LA MODELISATION EN BIOLOGIE

Abstract

We describe the structure of EDORA, an expert system which enables biologist to build his model. Our goal is to provide a system which is easy to use and which helps the biologist 'transform' their view of the problem from biological terms to a mathematical model. The elementary model permits users to construct models step by step. A set of commands provides guidance about how to construct appropriate functional scheme. This approach provides a framework for guiding the user during the process of model construction.

Résumé

Nous décrivons EDORA, un système expert d'aide à la modélisation en biologie. Notre but est de construire un système expert facile d'accès, aidant le biologiste à traduire en termes mathématiques ses données biologiques. La notion de modèle élémentaire permet à l'utilisateur de construire son modèle pas à pas. Un ensemble de commandes est prévu afin de le guider lors de la construction de son schéma fonctionnel. L'objectif principal de la maquette est d'aider l'utilisateur dans la construction de son modèle, en lui proposant une méthodologie de modélisation.

Contents

1 Introduction

The field of Artificial Intelligence, and especially expert systems is still experimental . Indeed there is no methodology for building knowledge bases. Biological modeling is perhaps not the easiest application domain for an expert system. This paper is devoted to such an application which has been recently worked out (EDORA meeting 88) see [16], in which the expert system has been developed with the aid of the SHIRKA software. The first section introduces SHIRKA while the second section deals with biological modeling. The last section describes more concrete aspects of the implementation, especially the structure of the knowledge bases.

2 SHIRKA

2.1 Notions about objects

In the classical relational formalism a production rule based system, an object does not exist in itself but only takes part in a set of statements spread throughout the knowledge bases. The reasoning task consists in handling those statements in order to deduce new ones. Objects are not represented explicitly; instead, they are described through properties spread throughout the rules. On the contrary, in the object formalism, the entities really exist and have their own space in the knowledge bases. Representations in structured objects stem from different ideas. They draw their inspiration from Minsky frames, Schank scripts and object-oriented languages. A frame is a structured representation of knowledge.

```
(frame-name
      (slot-1  (facet-11  value-11)
               (facet-12  value-12)
                      ....
               (facet-1n  value-1n))
             .........
      (slot-p  (facet-p1  value-p1)
                      ....
               (facet-pm  value-pm)))
```

A frame defines a class of objects (a class frame), a particular object is also described by a frame (an instance).

2.2 Introduction to SHIRKA

SHIRKA makes use of an object-centered knowledge representation model see [22]. Knowledge is entirely coded in a class and its representatives called instances. A class and its instances have attributes that represent a property, a characteristic or a connection to other classes or instances. An attribute in turn, is described, by a set of slots; each slot contains either a frame for a class or a frame for an instance. SHIRKA allows the user to define classes of interdependent objects, to create representatives of those classes, and to ask their respective attributes. SHIRKA tries to answer to those questions by using the introduced knowledge as also inference mechanisms. The basic inference mechanisms attached to an object-centered representation are: instanciation, inheritance, pattern-matching and procedural inference.

2.2.1 Description of functionalities

A facets is a predefined primitive; the following lists those currently available :

Facets of type and restrictive conditions :
$one, $list-of, $domain, $interval, $verify, $except, $card-min, $card-max.

Facets for slot value determination :

- $value defines the value of a slot in a frame common to all objects of the corresponding class, or a description which can be used to obtain this value by pattern matching.
- $if-necessary permits the association of computation and makes possible a procedural attachment. The corresponding methods are also described by frame; they are generally written in algorithmic language.
- $default ; a value is assigned to a slot if all other ways failed.

Facets for handling variables :
$var<-, $var->, $var-list<-, $var-list->, $var-name.

Facets of reaction in event (deamons) : $before-add, $after-add, $before-modif, $after-modif, $before-sup, $after-sup, $if-success, $if-check

Facet of commentary : $com

2.3 Inference mechanisms and reasoning

- Instanciation :
  This is the fundamental inference mechanism. It corresponds to the completion of a frame, by obtaining values of slots of the corresponding class and also values of slots inherited from other class frames.

- Inheritance :
  If the value of a slot is defined at a level, all the instances of this class and of more specific classes inherit this value.

- Pattern-matching :
  The facet $value in a class frame can be followed, by a list of patterns which are frame specializations of existing frames : they are descriptions of instances which can be assigned to slot value. The pattern matching consists of finding this slot value.

- Procedural inference : The value of a slot can be computed by an external algorithmic routine. These routines are themselves described by frames.

- The management of coherence :
  The management of coherence between instances is realize through the routines connected to facets and activated during manipulation of their values like adding, suppressing, modify.

- The classification :
  This allows the placement of an instance in a hierarchy, i.e., to determine automatically the instance of the most specialized frame which corresponds to the description.

See [1] for more details.

2.4 Conclusion

This introduction to SHIRKA shows that it corresponds to the requirement of the EDORA user. Indeed, SHIRKA offers in comparison with similar systems, specificities well adapted to our problem : procedural inference, pattern-matching and an homogeneous representation of knowledge.

3 Modeling in biology

Modeling in biology is not the easiest area to consider for the building of an expert system for modeling assistance, see [13]. Formalization and handling of knowledge depend on the know-how of the expert and on the needs and constraints he expresses. Our aim is to observe a biologist while he is formalizing a problem in order to fix the different steps and then to deduce the structural objects of the knowledge base suitable to biological modeling.

3.1 Models in biology

From experimental results and a priori knowledge, the biologist expresses a hypothesis which accounts for a biological phenomenon. So he builds a model, i.e., a simplified and incomplete representation of the actual phenomenon. Then to insure that the representation is realistic, the biologist validates the model by returning to the experimental results, followed by the appropriate modifications. Modeling provides a theoretical base for reasoning, which may lead to modifying the model and hence to producing new hypotheses. Mathematical models are often represented by system of ordinary differential equations system (we restrict aourselves to such models), the integration of which gives the dynamic behavior of the variables. Changing conditions or new hypotheses lead to an evolution of the mathematical expression. The modeling process can be divided into three stages

1. the first one consists of a characterization of the system, apart from its environment.

2. then comes the definition of characteristic variables in order to propose a functional scheme of the system, from which the mathematical model will be derived. The more complex the studied system is, the more difficult will be this stage.

3. the last stage is the definition of the mathematical model.

Note that the functional scheme representation is of great interest as an intermediate tool before the mathematical formulation see [19]. Indeed this representation allows one to fix a role for every variable and every interaction between variables. The mathematical modeling permits one to account for the studied phenomenon but oversimplifies it by showing only a partial view of the system. Therefore it's very important to propose tools that can manage all the knowledge in the given area.

4 The EDORA expert system

A system expert approach in a new domain imposes a plan of work. Schematically the resolution of a problem in Artificial Intelligence requires three stages.

1. understanding of the problem
2. choice of a convenient representation of the constituents
3. definition of a suitable resolution method.

These three stages must be processed together. The first stage, also called conceptualization consists of a transfer of expertise to build an expert system. The problem is to fix conceptual categories necessary for setting up a knowledge base, and to define static and dynamic links between these concepts. The expert makes successives tries and must be able to change the previous model at all time.

4.1 Important aspects

Every object requires a reference to the working context and a link to the other base constituent. Different description levels are distinguished. A representation based on objects allows to unify the representation used in different steps, to introduce a hierarchy and to use the inferences. The definition of data suitable for modeling is a fundamental aspect of the work. The system can store and use the heuristics and knowledge specific to the modeled domain. EDORA provides the biologist with a methodology for mathematical modeling from a biological phenomenon described by a biologist.

4.2 Functionalities

Knowledge modeling in biology allows the building of objects distributed among the different level of global system apprehension see [9].

- Input level :
  This is usually the semantic level in the system. The user defines the biologic objects to handle, with their characteristics relevant to the working situation. He then defines the relationships between these objects (for example : a species is a predator for another one). More information can be provided on the relationships themselves or the global situation composed of objects and their relationships.

- Mathematical modeling level :
  From the given biological situation, the user proposes a mathematical representation expressed by equations. The system itself can suggest a model selected from a predefined catalogue. The system checks the consistency between the equations and the initial biological situation and may query the user in order to maintain this consistency.

- Treatment level :
  The system computes the required mathematical treatments. The mathematical treatment that can be used include :

  - Identification
  - Graphic visualization of a variable in accordance with another one
  - Numerical integration
  - Study of stationary points and their stability
  - Periodic solutions search
  - Formal calculus

All these objects carry a semantic aspect that helps in using the expertise. The semantics are context dependent. The assignment of a specific sense to an object derives from the situation in which it will be used. In the EDORA system, the semantic aspect consists in storing a working environment pointing to other objects. The most important information is the links between

objects. For example, a mathematical variable includes in its object representation a reference to the biological object and to the physical object it model. The effective separation between the input level and the modeling level can be argued as follows :

1. biologists reason better at semantic level than at a mathematical one
2. this separation provides a precise building methodology that distinguishs a biological object and its mathematical representation.
3. it pushes biologists into developing new methods based especially on relationships between objects. That can result in new kinds of reasoning.

4.3 Conceptualization

The different steps in biological modeling and previous arguments lead us to define more precisely the object structure and the structure of the knowledge base itself. We define three modeling oriented objects.

Elementary objects :
there are two types of elementary objects

- Variables : A variable is a basic object, for example : bacteria
- Processes : A process describes an observed phenomenon, for example : growing-up, nutrition

Elementary model objects :
They represent possible relationships between elementary objects. An elementary model object is composed of one or two variables (or sometimes more)and a specific process.

Situation objects :
A situation object describes a specific working situation. It is composed of elementary model objects.

5 The prototype 88

5.1 Description

The aim of this prototype is to guide the user in the study and the description of the biological phenomenon. The main goal of the system is to help the user to build his model at a biological level and then at a mathematical one, by proposing a modeling methodology. We will present you the overall approach and the individual steps. The idea which emphasizes the knowledge base organization and the intelligent system approach is as follows :
The purpose is to describe a particular work situation using elementary models. An elementary model is composed of one or two biological variables and a specific process defining the type of interaction between those variables. A situation object is composed of a list of elementary models. This is a novel approach for a biologist, which is why we need commands suitable for guiding the user which is building his functional scheme. As a matter of fact, as we determined by experiment, biologist reasons using hypotheses that are difficult for him to make explicit. In order to determine the different objects used by the biologist, we propose the following approach :

- first of all, build a basis from an elementary model composed of the principal variables and a process that shows the interaction between those variables.
- introduce step by step other elementary models that are connected to the basis elementary model, up to the global description of the given biological phenomenon (called the situation).

5.2 Commands

During modeling, the biologist is acquainted with at least one of the variables that occurs in his experiment. The biologist tries to define the kinds of actions that affect his variables and the interactions between his variables and the other ones.

5.2.1 PATH commands

The aim of this command is to help the user in constructing his situation object. If the user handles two variables and he is trying to define the relationship between those variables, there are two possibilities :

- They can be directly connected by one process; that means that there is an elementary model including those two variables.
- They do not have a direct influence. In that case, the PATH command gives the all possible paths from one variable to the other one.

This command enables us to determine the list of possible elementary models that may describe the biological phenomenon. This kind of result can guide the user in building his situation object.

5.2.2 ANALYZE-MOD command

We assume the user builds his basis elementary model, then searches for other elementary models in the knowledge base connected to the first one, That is, those elementary models that include at least one of the main variables (corresponding to the basis elementary model variables). This command provides all the elementary models connected to the basic elementary model. ANALYZE-MOD and PATH guide the user to build the object situation which models the biological phenomenon.

5.2.3 CREATE-MOD-ELE command

This command creates an instance of the elementary model scheme in an interactive manner. The user must provide the biological variables and the process name.

5.2.4 CREATE-SITUATION command

This command interactively creates an instance of situation scheme using the elementary models chosen by the user.

5.2.5 CREATE-MOD-MATH command

Every biological elementary model is connected to a mathematical model. The latter is defined by : variables, parameters and process velocity. The user chooses a mathematical representation of all the biological objects he handles :

1. Variables of the mathematical model is the mathematical representation of the associated biological variables.
2. The process velocity is the mathematical representation of the associated biological process (for instance, it can represent the probability of interaction of variables).
3. The mathematical model parameters represent the velocity constants of the process related to every variable.

From these data, we produce a system of differential equations. Note that these equations are formal objects of the knowledge base.

5.3 Example : The Lotka-Volterra system

An ecologist wants to model the evolution of the lynx and hare population in Alaska. He first builds the biological model : the object situation which will be composed by three elementary models :

- mod1b : predation of the hare by the lynx.
- mod2b : hare reproduction
- mod3b : lynx death

The biologist builds his model using three processes and two variables, then he choses the mathematical representation of his biological variables.

```
  x1 the hare
  x2 the lynx
mod1b -> mod1m x1' = - a1 * x1*x2
               x2' = a2*x1*x2
mod2b -> mod2m x1' = a3 * x1
mod3b -> mod3m x2' = -a4 * x2
```

The evolution of each variable is influenced only by the elementary models in which it occurs. When adding up all the effects we obtain the famous Lotka-Voltera model :

```
x1' = - a1*x1*x2 + a3*x1
x2' = a2*x1*x2 - a4*x2
```

remark : if the biologist want to take into account the hare death and the lynx reproduction, he must introduce two elementary models. Then we obtain :

```
x1' = -a1*x1*x2 + a3*x1 - a5*x1
x2' = a2*x1*x2  - a4*x2 + a6*x2
```

See [8] for more details.

5.4 The knowledge base architecture

The function of EDORA is to help the biologist when he is building his object situation. The knowledge base must be structured in such a way as to permit the user to describe and specify his situation, as precise and clear as possible. It is in this aspect that the classification is a valable assistant, because it structures and then makes usable the user data. The classification is well adapted to what we want to do, i.e., to guide a non-specialist in the topic. The system must take into account

- The pedagogical aspect of the expertise communication
- The diversity of the handled concepts
- Notion of a point of view

We divide the knowledge base into different layers corresponding to the different apprehension levels of the global system.

5.4.1 The input level

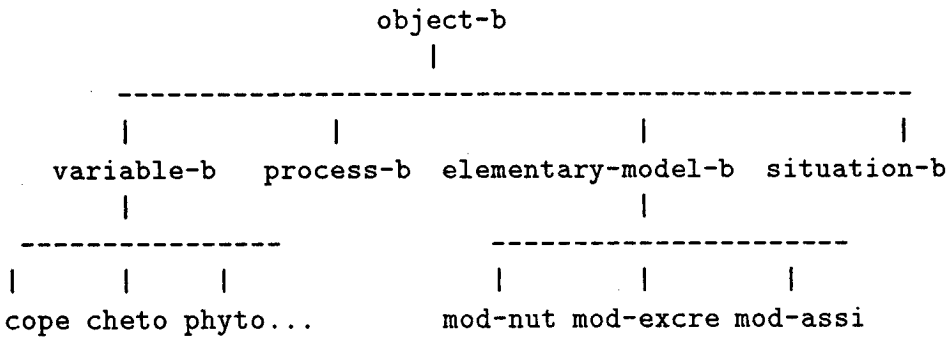
This corresponds to the system biological level. From the beginning we provide for many entrance points of the system. The user can describes his problem using different points of view.

- The phenomenological point of view corresponds to the description of the observed phenomenon, for example : growth, death...
- The point of view corresponding to variables
- The modeling point of view includes the observed phenomenon and variables.

All these points of view stand for a particular knowledge which will be separated in three classes of objects. These classes of objects will be represented in a hierarchy which the thread will be the problem description support. In order to simulate expert behavior, we accomplish the task into two steps :

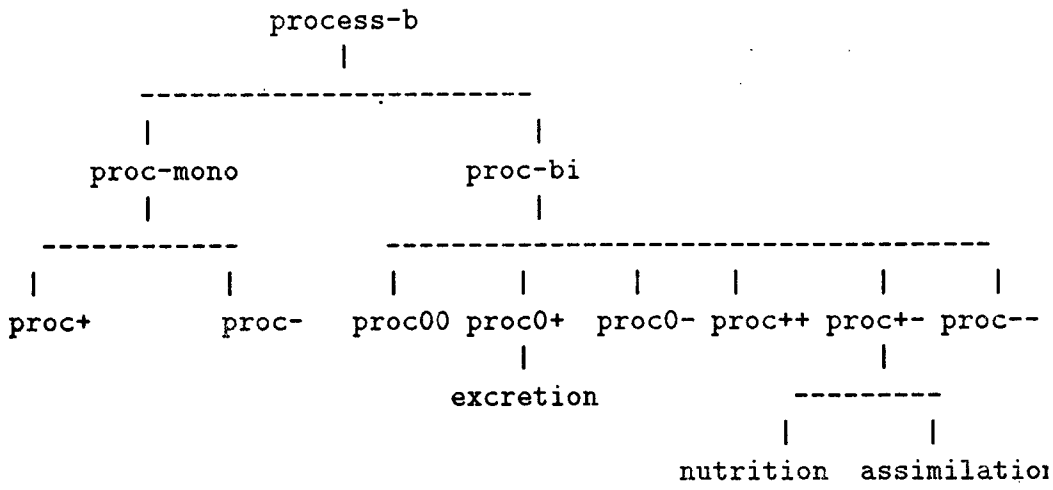
- First the user acquires information and tries to match them to his knowledge.
- Then he makes a synthesis by regrouping the knowledge.

The initial classification of variables and processes determines the attribute of an elementary model instance (this is the first step). The second classification classifies this instance in the elementary model class, and then permits access to the specific knowledge of this class.



In modeling, as we mentioned before, the biological process is the object most difficult to identify. In fact, the process defines the biological phenomenon. The classification of this class uses two criteria :

- The number of variables used by the process
- The action that the process can perform on a variable : nothing, increment, decrement (for instance : the process of lynx death decresses the number of lynxes).



5.4.2 Physical level

It is the knowledge level related to all the physical concepts like : unit, experimental point, measures, parameter. These objects will be connected to the biological data we manipulate.

5.4.3 The mathematical-formal level

From the description of the mathematical model,i.e., variables parameters and process velocity, we build a formal object which is a differential equation. We use two programs. One displaying the formal object and the other one for reading it. The displaying requires three steps :

- the lisp object is transformed into a list in which the first argument is the operator and the remainder the operator arguments. Example : a/b becomes (/ a b)
- the list is transformed into a graphic object
- the graphic object is displayed.

The formal level evaluates the equation and creates the formal object associated with it. Note that we can transform a SHIRKA object, i.e., a scheme into a formal object and inversely. We can also convert the formal object into a fortran object; this enables us to use many programs like :

- numerical integration
- sensibility analysis
- study of the stationary points

6 Conclusion

We presented the general architecture of EDORA. The system offers a methodology for modeling in biology. We try to define structural objects of the knowledge base well adapted to modeling. Initially, the biologist builds his model and then tries to study it. We focused our effort on the first step, i.e., how to help the user in the building of his model. We also enable the handling of formal equations in order to connect to software for treatment. There are still many things to do in EDORA. First, building a graphics interface, then, creating a dialogue environment for important commands, i.e., building a command hierarchy, for better guiding the user. We must also provide for explanation at every level of system use, and extend the system's applicability in analyzing expert behavior.

Artificial intelligence, in particular expert systems, is increasingly used in the biological-ecological and bio-medical areas; EDORA is only one example of such research . But it points out that AI has more then simply a practical aspects : it stimulates research, it raises questions, it creates ideas. The present version of EDORA is not a commercial product. Our aims consist of producing successive prototypes which only are partial realizations. These prototypes raise fundamental questions and stimulate the progress in biology and A.I. From this point of view, we consider that EDORA has already given results. The EDORA realization has led us to consider many new concepts such as processes, elementary models and functional scheme. SHIRKA was built for EDORA, and is an original, complete, efficient tool, perfectly adapted to needs of the EDORA user.

References

[1] J.L AGUIRE CERVANTES, D. BLOCH , F. RECHENMANN, N. ROUIBAH 87. SHIRKA : compilation, explication et cohérence dans les bases de connaissances centrées-objet. MARI 87, Paris, 1987.

[2] V ANDERSEN 82. Mise au point d'un modèle d'écosysteme planctonique en mer Ligure. DEA, Université Pierre et Marie Curie, 1982.

[3] C. BINOT, D. ROGER 87. Présentation graphique d'informations : aspects informationnels. Cognitiva 87.

[4] G. CHAILLOUX 85. Le-Lisp version 15.2. Manuel de référence.

[5] A. CHERUY 88. Méthodologie de la modélisation. In Les cahiers d'EDORA, Rapport de recherche INRIA 866, A.PAVE, 1988.

[6] N. GAUZIN 87. Le projet EDORA : un exemple de système à base de connaissance en biologie. DEA, Université Lyon 1, 1987.

[7] J.G. GANASCIA 85. La conception des systèmes experts. La recherche, oct.85, 170.

[8] J.L. GOUZE 88. Structure et comportement des modèles mathématiques en biologie. Rapport INRIA 869, 1988.

[9] J.L. GOUZE, P. VIGNARD 85. EDORA : un système intelligent d'aide à la modélisation en biologie. AMSE Review, volume 2, n 5, 1985.

[10] M.K. HAMROUNI 79. Etude et développement d'un système informatique d'aide à l'élaboration de modèles en biologie. Thèse de 3 cycle, Université Paris 6, 1979.

[11] M.C. HATON, B. DEVIN, L. PIERRON, P. CAUBERE 87. Représentation des connaissances et apprentissage en synthèse organique : une démarche experte. Cognitiva 87.

[12] D. HERIN-AIME, O. MASSIOT 87. DEMSI : un prototype de système expert orienté objet pour l'évolution des systèmes d'information. Cognitiva 87.

[13] J.D. LEBRETON, C. MILLER 82. Modèles dynamiques déterministes en biologie. Masson, 1982.

[14] R.W. LINGARD 81. A software methodology for building interactive tools. Conf on software engin march 9-12 1981 San Diego CA.

[15] N. PACAULT 87. Elaboration d'un système d'aide à la modélisation en biologie. DESS,EHEI.

[16] A. PAVE 88. Les cahiers d'EDORA, Rapport INRIA 866, juillet 88.

[17] A. Pave 87. Interprétation et construction de modèles de la dynamique des populations à l'aide de schémas fonctionnels. In Les cahiers d'EDORA, Rapport INRIA 866, juillet 88.

[18] A. PAVE, F. RECHENMANN 86. Computer-aided modelling in biology : an Artificial Intelligence approach. In "A.I.Applied to simulation", Ed.Kerckhoffs, Vansteenkiste G.C., Zeigler B.P.. SCS Simul.Seri, 18, 1986, 52-66.

[19] A. PAVE 87. Systèmes experts : application à la modélisation en biologie. MATAPLI,11,p,1987.

[20] C. PIERRET 88. Structuration des connaissances et raisonnement à l'aide d'objets. Rapport INRIA 847, mai 88.

[21] F RECHENMANN, A PAVE 85. Object Oriented Knowledge Representation. Intern. Working Conf. on Artificial Intelligence in Simulation,Gand,1985.

[22] F RECHENMANN 87. SHIRKA : Système de gestion de base de connaissances centrées-objet. Manuel d'utilisation. INRIA, 1987.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique